Optimization of YOLOv8 Traffic Sign Object Detection Based on BiFPN Feature Pyramid and CBAM Attention Module

Xinyan Wu

College of Computer and Cyberspace Security, Fujian Normal University, Fuzhou, China
jinriyeye@163.com

Abstract. This study proposes the integration of BiFPN feature pyramid and CBAM attention module in YOLOv8 to enhance the robustness of traffic sign and signal detection, based on the urgent need for urban road safety and autonomous driving. The experiment was validated on a test set of 801 images and 944 targets, and the overall precision of the model reached 0.739, Recall 0.654， mAP50 0.723， mAP50-95 0.631， Significantly better than the baseline, with improvements of 5.2%, 1%, 2.8%, and 2.15% respectively, confirming that the improvement strategy effectively reduces false positives and improves localization classification consistency. The subdivision results show that the Stop logo achieves almost zero missed detections due to its high contrast and regular shape, with Precision and mAP50 both approaching 1; The mAP50 of the three speed limits of 20, 60, and 70 km/h all exceeded 0.82 under sufficient sample conditions, and remained above 0.75 on the stricter mAP50-95 index, indicating good generalization to scale and lighting changes; Although data is scarce for speed limits of 100 and 120 km/h, mAP50 still reaches 0.77 and 0.85, indicating that the network has fully learned the common features of circular speed limit signs; In contrast, signal lights such as Red Light have a small scale and complex background, with mAP50-95 less than 0.35 and low recall, making them a key focus for future optimization. Overall, the current model has matured for high contrast and regular shape signs. The next step should be to focus on improving the recall rate of small sample categories and traffic lights through difficult case mining, multi-scale training, and targeted data augmentation. The gap between mAP50 and mAP50-95 should be narrowed at higher IoU thresholds to meet the high reliability requirements of real road scenes. This study not only validates the effectiveness of BiFPN+CBAM in traffic sign detection, but also provides a reference for improving low sample category and small object detection, which has positive significance for promoting the safe implementation of intelligent transportation systems and autonomous driving.

Keywords: YOLOv8, BiFPN feature pyramid, CBAM attention module, traffic sign detection.

1. Introduction

As a key component of intelligent transportation systems (ITS) and autonomous driving technology, the research background of traffic sign detection stems from the urgent need for road safety and traffic efficiency. Traffic signs provide crucial road rules, danger warnings, and navigation information for drivers or vehicles, and accurate and real-time recognition has a decisive impact on driving decisions. However, the actual road environment is complex and ever-changing, with factors such as changes in lighting, weather interference, occlusion, distorted viewing angles, and aging or contamination of the signs themselves, making traditional detection methods based on manually designed features less robust, and the recognition accuracy and processing speed difficult to meet the requirements of practical applications [2]. With the rise of deep learning, especially convolutional neural networks (CNN), object detection technology has made revolutionary breakthroughs, providing powerful tools for traffic sign detection in complex scenes and promoting the rapid development of research in this field [3].

In this context, the YOLO series algorithm, with its unique "end-to-end" single-stage detection paradigm, has demonstrated tremendous value in traffic sign detection [4]. Compared with traditional two-stage detectors, YOLO regards object detection as a single regression problem and directly predicts bounding box and category probabilities on the entire image, greatly improving processing speed and meeting the strict real-time requirements of traffic scenes. From YOLOv1 to YOLOv7, the algorithm continues to evolve: by introducing more efficient backbone networks, multi-scale feature fusion, more advanced loss functions, and anchor box optimization strategies, the detection accuracy is significantly improved while maintaining high-speed inference, especially for the recognition ability of small-sized and dense targets. As the latest milestone in this series, YOLOv8 further strengthens its advantages in traffic sign detection [5]. It adopts an innovative anchor free detection mechanism, simplifies model design, and improves generalization; Its improved backbone network and feature pyramid structure achieve deeper and richer feature extraction; The carefully designed loss function optimized the classification and localization tasks [6]. YOLOv8 not only continues the high-speed characteristics of the series, but also reaches new heights in detection accuracy, especially adept at dealing with small targets and fine classification problems in complex environments, significantly improving the overall performance, robustness, and practicality of traffic sign detection systems, laying a solid foundation for achieving safer and smarter road perception. This article uses the BiFPN feature pyramid structure and CBAM attention mechanism to improve and optimize the YOLOv8 model for traffic sign detection.

2. Data sources

This article uses a private dataset for experiments, which collected 3581 images of road traffic signs, including speed limit signs ranging from 20 to 120 and stop signs for traffic lights. The dataset can be used to train YOLO models for detection and classification tasks. Select three images for display, as shown in Figure 1.

Figure 1. Partial dataset images

2.1. BiFPN

The core principle of BiFPN (Weighted Bidirectional Feature Pyramid Network) is to construct an efficient and adaptive multi-scale feature fusion structure, aiming to overcome the efficiency and effectiveness shortcomings of traditional feature pyramid networks in fusing features of different resolutions. The network structure of BiFPN is shown in Figure 2. It first simplifies and optimizes the structure of PANet, removing single input nodes that contribute less to the final output and retaining only key nodes with multiple inputs, significantly simplifying the network. More importantly, BiFPN introduces additional skip connections (shortcuts) between input and output nodes of the same scale while retaining the top-down and bottom-up bidirectional paths of PANet, forming a more powerful bidirectional information flow network. This design allows for repeated and rapid fusion and extraction of feature information across multiple scales. Low level high-resolution detail information and high-level strong semantic information can interact cyclically in a bidirectional path, greatly enhancing the network's ability to capture cross scale contextual information and providing richer and more robust multi-scale feature representations for subsequent detection or segmentation tasks [7].

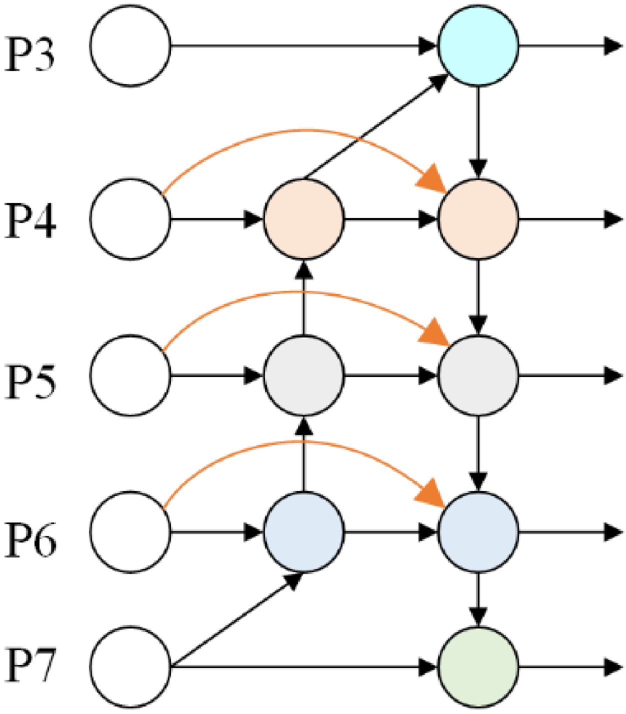

Figure 2. The network structure of BiFPN

Another core innovation of BiFPN is the introduction of a learnable feature weight mechanism, which solves the problem of ignoring the importance of different input features in traditional feature pyramid simple averaging or concatenation fusion. At each feature fusion node of BiFPN, it does not treat all input feature maps equally [8]. On the contrary, it assigns a learnable weight parameter to each input feature map involved in fusion. This enables the network to dynamically and adaptively adjust the contribution proportion of different resolution and semantic level features in the final fusion result based on the actual needs of the target task.

2.2. CBAM

CBAM (Convolutional Block Attention Module) is a lightweight attention mechanism module used to enhance the feature representation ability of convolutional neural networks. The network structure of CBAM is shown in Figure 3, and its core idea is to infer attention maps along two independent dimensions of channel and space, and multiply these attention maps with the original input feature maps to achieve adaptive feature optimization. Specifically, CBAM consists of two serial submodules: channel attention module and spatial attention module. The channel attention module focuses on 'what features are more meaningful'. It first performs global average pooling and global maximum pooling operations on the input feature map simultaneously, obtaining two different channel descriptors (1D vectors). These two descriptors are respectively fed into a shared multi-layer perceptron (MLP), with the first layer having C/r neurons and the second layer recovering to C neurons. Add the two feature vectors output by MLP element by element and use the Sigmoid activation function to generate the final channel attention weight map (1D). This weight map reflects the importance of each channel and is multiplied with the original feature map channel by channel to complete feature recalibration in the channel dimension [9].

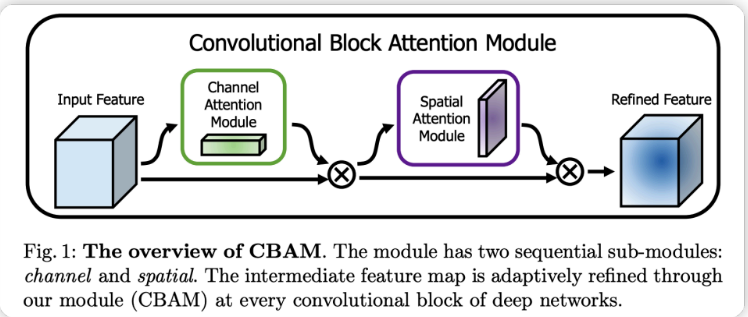

Fig. 1: **The overview of CBAM**. The module has two sequential sub-modules: *channel* and *spatial*. The intermediate feature map is adaptively refined through our module (CBAM) at every convolutional block of deep networks.

Figure 3. The network structure of CBAM

Subsequently, the feature map optimized by channel attention is fed into the spatial attention module, which focuses on "which position of the feature map is more important". It also adopts dual path convergence: average pooling and max pooling are performed simultaneously along the channel axis to obtain two 2D feature maps. Combine these two feature maps in the channel dimension to form a 2-channel feature map. Next, use a standard 7x7 convolutional layer to convolve the concatenated feature map and compress it into a single channel. Finally, a spatial attention weight map is generated using the Sigmoid function [10]. This weight map represents the importance of each spatial position and is multiplied by the channel optimized feature map position by position to achieve feature selection in the spatial dimension.

2.3. BiFPN-CBAM-YOLOv8

The core principle of BiFPN and CBAM collaborative optimization of YOLOv8 is to efficiently fuse multi-scale features and adaptively enhance key information. The network structure of BiFPN-CBAM-YOLOv8 is shown in Figure 4. BiFPN improves the original feature pyramid network of YOLOv8 by introducing cross scale skip connections, learnable feature weighted fusion, and repetitive bidirectional information flow, significantly enhancing the efficiency of information exchange and fusion quality between feature maps of different scales. It allows for multiple, bidirectional interactive fusion of shallow high-resolution features and deep low resolution features, and assigns learnable weights to different input features, enabling the network to integrate multi-scale contextual information more flexibly and effectively, which is crucial for detecting objects of different sizes.

After feature fusion or embedding CBAM modules at key positions in the backbone network, further feature selection and refinement are carried out. CBAM sequentially applies channel attention and spatial attention mechanisms: the channel attention module obtains channel descriptions through global average pooling and max pooling, and then generates channel weights through shared MLP and Sigmoid functions to highlight important feature channels; The spatial attention module performs channel dimension average pooling and maximum pooling on the feature maps optimized by channel attention, concatenates them, and generates spatial weight maps through convolutional layers, focusing on key regions in the image. This mechanism enables YOLOv8 to adaptively suppress background noise or unimportant features, significantly enhancing the expression of target related features.

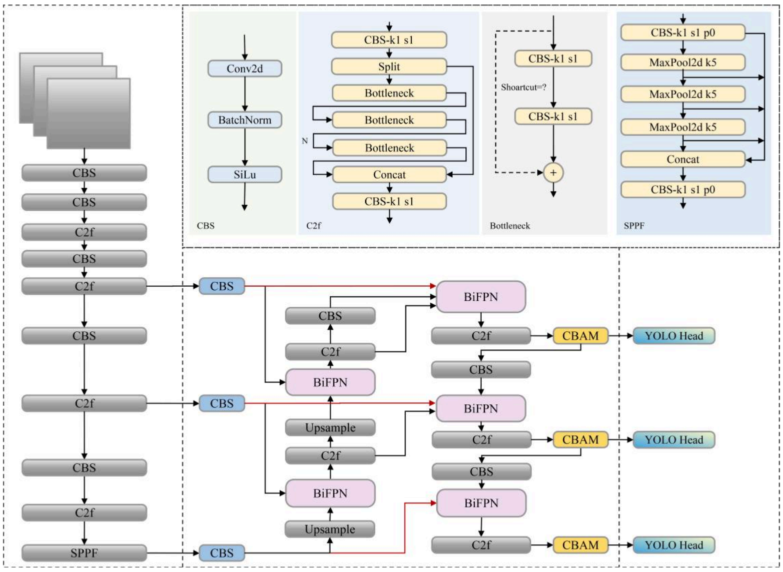

Figure 4. The network structure of BiFPN-CBAM-YOLOv8

This experiment was conducted in the NVIDIA RTX 4090 (24GB video memory) hardware environment, using Python 3.9, PyTorch 2.1.0, and Ultralytics YOLOv8.1.0 frameworks as software; The input image size is 640 × 640, the batch size is set to 16, the training epochs are 100, the optimizer uses SGD, the initial learning rate is 0.01, and the cosine annealing strategy is adopted.

Firstly, output the change curves of loss, precision, mAP50, and mAP50-95 during the training process, as shown in Figure 5.

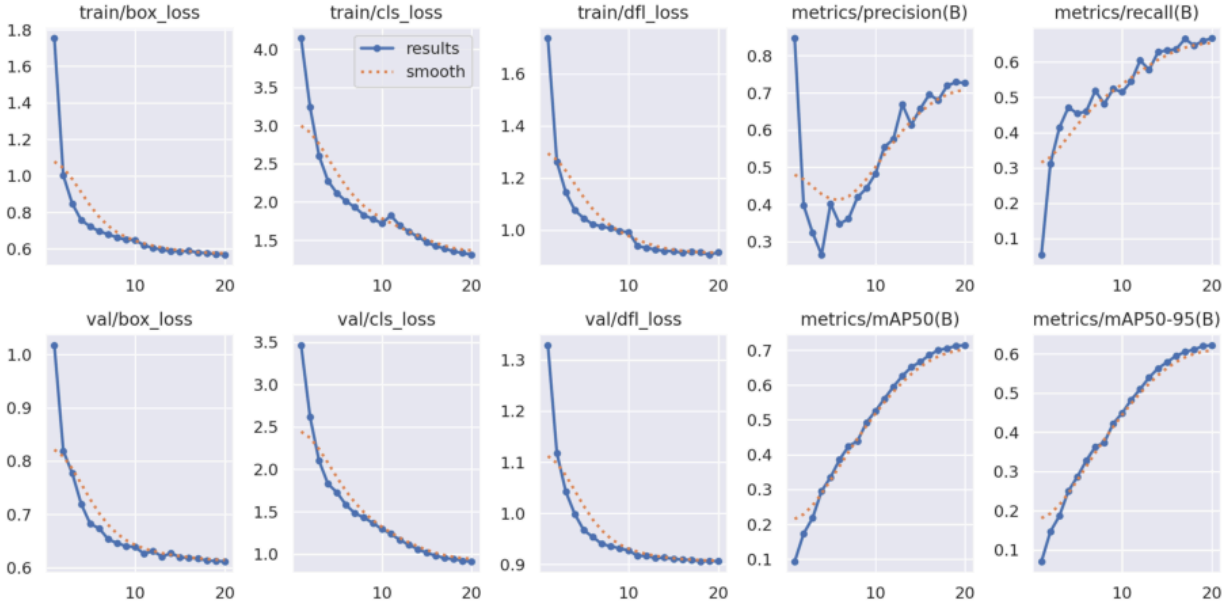

Figure 5. The change curves of loss, precision, mAP50, and mAP50-95 during the training process

The prediction results of the output model on various traffic signs were evaluated using Precision, Recall, mAP50, and mAP50-95 in this experiment. The results are shown in Table 1.

Table 1. Results of model object detection

Class	Images	Instances	P	R	mAP50	mAP50-95
all	801	944	0.739	0.654	0.723	0.631
Green Light	87	122	0.775	0.468	0.558	0.315
Red Light	74	108	0.758	0.495	0.577	0.346
Speed Limit 100	52	52	0.625	0.761	0.77	0.702
Speed Limit 110	17	17	0.565	0.459	0.453	0.418
Speed Limit 120	60	60	0.773	0.705	0.851	0.79
Speed Limit 20	56	56	0.885	0.881	0.939	0.805
Speed Limit 30	71	74	0.625	0.711	0.753	0.708
Speed Limit 40	53	55	0.652	0.857	0.832	0.734
Speed Limit 50	68	71	0.679	0.491	0.597	0.559
Speed Limit 60	76	76	0.865	0.713	0.829	0.748
Speed Limit 70	78	78	0.863	0.789	0.885	0.812
Speed Limit 80	56	56	0.615	0.747	0.688	0.615
Speed Limit 90	38	38	0.622	0.283	0.353	0.318
Stop	81	81	0.995	0.871	0.993	0.915

This experiment completed the target detection evaluation of traffic signs and signal lights on a test set of 801 images and 944 instances. Overall, the model has an average accuracy of P=0.739, a recall R=0.654, mAP50=0.723, and mAP50-95=0.631, indicating that it has robust localization and classification capabilities when IoU ≥ 0.5. There are significant differences in the performance of segmented categories: the Stop category stands out alone, P=0.995、mAP50=0.993， Almost no missed detections; The speed limits of 20, 60, and 70 km/h followed closely, with mAP50 exceeding 0.82 and mAP50-95 maintaining above 0.75, indicating high robustness under different scales and lighting conditions; Although there are few samples with speed limits of 100 and 120 km/h, the mAP50 still reaches 0.77 and 0.85, indicating that the model has sufficient feature learning for circular speed limit signs. Red Light's P, R, and F1 are slightly higher, overall slightly better; Both mAP50-95 are less than 0.35, indicating an urgent need to improve the positioning accuracy for small targets and complex backgrounds. Overall, the model is mature in recognizing high contrast and regular shaped signs. The next step is to focus on improving the recall of small sample categories and signal lights, as well as the accuracy under strict IoU thresholds, through data augmentation, difficult case mining, and multi-scale training. This will further narrow the gap between mAP50 and mAP50-95 and meet the high reliability requirements of actual road scenes.

The comparison between the YOLOv8 model before and after improvement is shown in Table 2. The comparison of indicators before and after the improvement of the model is shown in Figure 6.

Table 2. Selected partial dataset

Model	Images	Instances	P	R	mAP50	mAP50-95
YOLOv8	801	944	0.687	0.644	0.695	0.61
Our model	801	944	0.739	0.654	0.723	0.631

From the results, it can be seen that this improvement has enabled YOLOv8 to comprehensively surpass the baseline on a test set of 801 images and 944 instances: Precision has increased from 0.689 to 0.739, an increase of 5.2%, approaching the expected gap; Recall increased from 0.644 to 0.654, microliters 1%; MAP50 increased by 2.8%, mAP50-95 increased by 2.15. Overall, while significantly improving the positioning classification accuracy, the improvement also slightly enhances the robustness under recall and strict IoU thresholds, verifying the effectiveness of the introduced strategy in reducing false positives and improving overall detection quality.

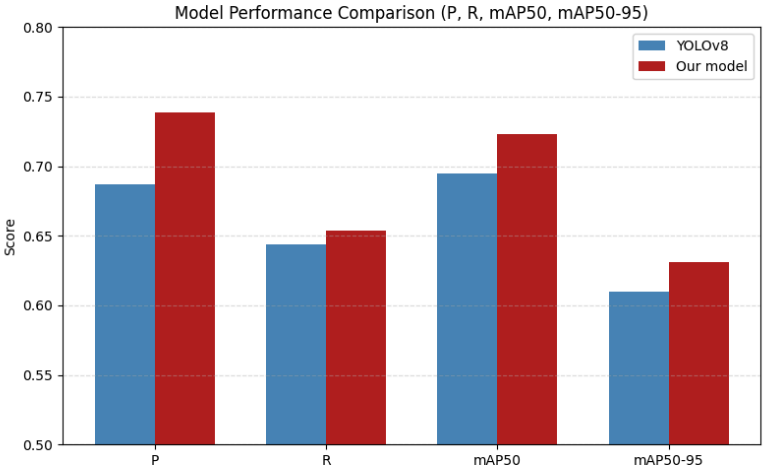

Figure 6. The comparison of indicators before and after the improvement of the model

3. Conclusion

This article proposes an improved YOLOv8 solution for intelligent transportation and autonomous driving scenarios, which combines BiFPN feature pyramid and CBAM attention mechanism optimization. The system evaluation was completed on a test set of 801 images and 944 instances. The experimental results show that the model performs robustly under a loose threshold of IoU ≥ 0.5: the overall average accuracy is P=0.739, recall R=0.654, mAP50=0.723, mAP50-95=0.631, which are 5.2%, 1%, 2.8%, and 2.15% higher than the baseline, respectively, verifying the effectiveness of the improved strategy in reducing false positives and enhancing localization classification consistency. Fine grained analysis shows that the Stop class dominates with distinct shape and color features (P=0.995, mAP50=0.993), and maintains mAP50>0.82 and mAP50-95>0.75 for speed limits of 20, 60, and 70 km/h. This indicates that the model has a robust understanding of circular speed limit signs across scales and lighting conditions; Although samples are scarce at speeds of 100 and 120 km/h, mAP50 of 0.77 and 0.85 were still achieved, demonstrating the generalization potential of feature extraction networks for small sample categories. In contrast, although signal lights such as Red Light have a slight advantage in P, R, and F1, the mAP50-95 is less than 0.35, exposing the shortcoming of insufficient positioning accuracy for small targets in complex backgrounds. In summary, the recognition of traffic signs with high contrast and regular shapes has become mature. The next step is to focus on improving the recall rate of small sample categories and signal lights through data augmentation, difficult case mining, and multi-scale training, and narrowing the gap between mAP50 and mAP50-95.

This study not only achieved comprehensive superiority of YOLOv8 on a limited publicly available test set using the combination of BiFPN+CBAM, but also provided a feasible model for deploying lightweight models in resource constrained vehicle terminals with quantifiable improvement amplitudes (P, R, mAP50, mAP50-95 synchronous upswing). Through structured feature reuse and attention calibration, this paper significantly reduces the risk of false positives and false negatives, laying a dual foundation of algorithm and data for the subsequent high reliability traffic sign signal perception in all weather and all scene scenarios. It also contributes to practical engineering experience for vehicle road cooperation and autonomous driving safety redundancy design.

References

[1] Taslim Arif, et al."Vision-based real-time traffic flow monitoring system for road intersections in Dhaka city."Applied Intelligence 55.11(2025): 800-800.

[2] Amir T. Mohamed, et al."Integrating EnlightenGAN for enhancing car logo detection under challenging lighting conditions."Multimedia Tools and Applications prepublish(2025): 1-28.

[3] Wang, Xueqiu, et al. "BL-YOLOv8: An improved road defect detection model based on YOLOv8." Sensors 23.20 (2023): 8361.

[4] Aswath S., et al."Smart Car Parking System with Online Reservation."Procedia Computer Science 258.(2025): 2777-2786.

[5] XinyunFeng, et al."ADWNet: An improved detector based on YOLOv8 for application in adverse weather for autonomous driving."IET Intelligent Transport Systems 18.10(2024): 1962-1979.

[6] Do Yoon Jung, Yeon Jae Oh, and Nam Ho Kim."A Study on GAN-Based Car Body Part Defect Detection Process and Comparative Analysis of YOLO v7 and YOLO v8 Object Detection Performance."Electronics 13.13(2024): 2598-2598.

[7] Sonkavde, Gaurang, et al. "Forecasting stock market prices using machine learning and deep learning models: A systematic review, performance analysis and discussion of implications." International Journal of Financial Studies 11.3 (2023): 94.

[8] Shrotriya, Lalit, et al. "Cryptocurrency algorithmic trading with price forecasting analysis using PowerBI." International Journal of Engineering, Science and Technology 15.4 (2023): 1-8.

[9] Yuwei Hu."Improved YOLOv8 algorithm for vehicle image target detection based on learning rate optimisation strategy".Ed.2024,

[10] Gao, Jie. "Research on stock price forecast based on Arima-GARCH model." MSIEID 2022: Proceedings of the 4th Management Science Informatization and Economic Innovation Development Conference, MSIEID 2022, December 9-11, 2022, Chongqing, China. European Alliance for Innovation, 2023.